Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7284–7287, October 1988
Genetics

IS*T2*: An insertion sequence from *Thiobacillus ferrooxidans*

(repetitive sequences/chemolithoautotroph)

JAMES R. YATES*, RICHARD P. CUNNINGHAM†, AND DAVID S. HOLMES‡

*General Electric Company, Center for Research and Development, Schenectady, NY 12301; †Department of Biological Sciences, State University of New York at Albany, Albany, NY 12222; and ‡Department of Biology, Rensselaer Polytechnic Institute, Troy, NY 12180

Communicated by Ivar Giaever, May 23, 1988

ABSTRACT **The genome of *Thiobacillus ferrooxidans* (strain ATCC 19859) contains at least two families of repeated sequences, termed family 1 and 2. The nucleotide sequence of a family 2 member was determined. It is 1408 base pairs long and has structural features similar to those of insertion sequences (IS elements). Terminal inverted repeats 25 base pairs in length are present. These inverted repeats are imperfect and adjacent to target-site duplications 9 base pairs in length. Several open reading frames were detected (the longest was 888 base pairs). We have named this IS element-like sequence IS*T2*. The ends of a second example of IS*T2* were analyzed and compared to those of the first. The DNA sequences are identical and similarly sized target-site duplications are present.**

Repetitive DNA sequences are present in the genomes of many diverse prokaryotes (1–5). In some cases (4, 6, 7) these repeated sequences have been shown to be transposable elements called insertion sequences (IS elements). Most of the IS elements identified to date have been found in *Escherichia coli* and related enterobacteria (see ref. 8 for a review). However, the presence of IS-like sequences in other species implies that they may be a common occurrence in the genomes of prokaryotes. The nucleotide sequences of most IS elements share structural features including inverted terminal repeats, target-site duplications, and at least one open reading frame (ORF). IS elements do not carry selectable markers and it is thought that the ORF encodes a protein involved in transposition (i.e., a transposase) (8).

The Gram-negative, chemoautotrophic bacterium *Thiobacillus ferrooxidans* derives all the energy required for growth from the oxidation of ferrous iron and sulfur compounds (9). In addition to its unusual metabolism, *T. ferrooxidans* is an extreme acidophile and is resistant to many metal ions (10). We previously showed that the genome of this organism contains at least two families of repeated DNA sequences (5). Southern blot analysis revealed that one of these families, called family 2, had approximately 16 members. The members of family 2 were located primarily on the chromosome of this strain of *T. ferrooxidans*. A comparison of the restriction sites in two recombinant plasmids bearing family 2 members indicated that these repeated sequences were about 1 kilobase (kb) long (5).

We determined the nucleotide sequence of a DNA fragment containing a family 2 member.[§] Analysis of these data revealed the presence of inverted terminal repeats, target-site duplications, and three major ORFs. The resemblance to known IS elements leads us to propose the name IS*T2* for this family of repeated sequences.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Media. Construction and characterization of the two recombinant plasmids, pTf32 and pTf34, bearing family 2 members have been described (5). These plasmids were isolated from *E. coli* C600 grown in Luria broth as described (5). Single-stranded DNAs used as templates in sequencing reactions were isolated from *E. coli* JM101 grown in 2YT (which contains 16 g of Bacto-tryptone, 10 g of Bacto-yeast extract, and 5 g of NaCl per liter).

Recovery of DNA Fragments from Agarose Gels. Recombinant plasmids were digested with restriction endonucleases in accordance with the supplier's specifications, and the resultant fragments were separated by agarose gel electrophoresis. After staining with ethidium bromide, bands corresponding to the fragments of interest were excised. The gel slice was inserted into a dialysis-tubing bag containing 5 ml of electrophoresis buffer, and the DNA was extracted from the gel by electroelution (100 V for 1.5 hr). The fragments were chromatographically purified on a small DEAE-cellulose (Whatman DE52) column and concentrated by ethanol precipitation.

Subcloning of a Family 2 Member. The recombinant plasmid pTf32 (Fig. 1) was digested with *Eco*RV and the fragments were inserted into the *Sma* I site of the M13 vector um21 (International Biotechnologies, New Haven, CT). JM101 cells were transformed with these recombinant M13 phage by the procedure specified by the vendor. Phage with 2 kb of DNA inserted in both orientations were selected by restriction enzyme analysis of M13 replicative-form DNA [isolated by a modification of the rapid heating technique (11)]. Further restriction enzyme analysis confirmed the presence of family 2 members in the inserts.

Generation of Deletions for DNA Sequencing. Single-stranded templates were isolated from cells bearing the recombinant phage mentioned above by a modified polyethylene glycol (PEG) precipitation procedure (J. Salvo, personal communication). Plaques were picked and used to infect JM101 cells. After 4 hr of growth the cells were pelleted by centrifugation, and recombinant phage were precipitated from the supernatant fluid by the addition of PEG 6000 and NaCl. The precipitated phage were harvested by centrifugation and resuspended in 10 mM Tris, 8.0/0.1 mM EDTA/0.3 M sodium acetate. One extraction with phenol was performed and the single-stranded templates were concentrated by ethanol precipitation. Nested sets of deletions were created by using the Cyclone system (International Biotechnologies) according to the manufacturer's recommendations.

DNA Sequence Determination. Nucleotide sequences were determined by the dideoxy chain-termination method (12) using [α-[^{35}S]thio]dATP (Amersham). Reaction mixtures were applied to 8% polyacrylamide gels containing 8 M urea and electrophoresis was carried out at 1700 V for 4 hr. The gels were fixed in 10% acetic acid. Autoradiography was

Abbreviations: IS element, insertion sequence; ORF, open reading frame.

[§]The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03859).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

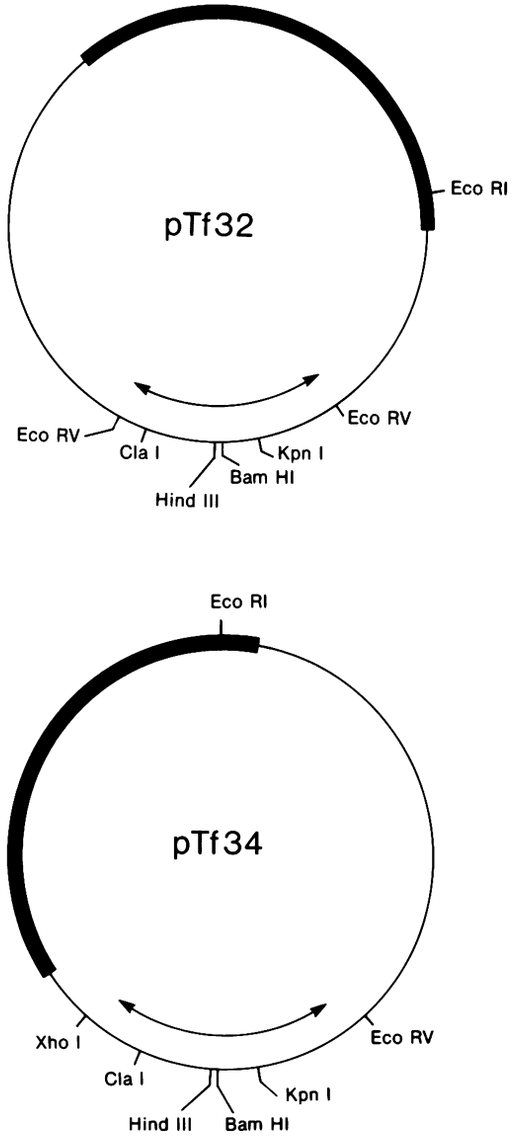

FIG. 1. Partial restriction maps of the recombinant plasmids used as sources of DNA fragments for sequence analysis. The thick black portion of the circles represents pBR322 (the position of the *Eco*RI site is indicated to help orient the insert). The inserted fragment of *T. ferrooxidans* DNA is shown as a thin line. Double-headed arrows indicate the approximate position of the family 2 member as determined by Southern blot analysis (5). More complete restriction maps of these two recombinant plasmids are presented in ref. 5.

accomplished by exposing dried gels to X-Omat AR film (Kodak).

RESULTS

Sequencing Strategy and Restriction Sites of IS*T2*. The 2-kb *Eco*RV–*Eco*RV fragment shown in Fig. 2 was selected for sequence analysis because previous data had indicated that the entire family 2 member was likely to be present on this fragment (5). Two recombinant plasmids (pTf32 and pTf34) with family 2 members showed conservation of sites between the *Kpn* I site and the closest *Cla* I site but showed no similarities outside this region. In addition, hybridization experiments in which fragments of one recombinant plasmid were probed with the other (5) indicated that the region of similar DNA sequences shared by pTf32 and pTf34 did not extend past the *Eco*RV sites indicated in Fig. 2*A*. The nucleotide sequence of one strand of this entire fragment was determined, but only the region from position 1 to 1560 was confirmed on the other strand. The arrows shown in Fig. 2*B* indicate the overlapping sequences obtained from several deletions for each strand. Some of the restriction sites derived from the DNA sequence of the region from position 1 to 1560 are indicated in Fig. 2*A*. The positions of several sites previously shown (5) to be within the family 2 member (i.e., *Cla* I, *Hin*dIII, *Bam*HI, and *Kpn* I) correlate well with the data obtained by mapping the recombinant plasmid.

Nucleotide Sequence of IS*T2*. The nucleotide sequence of a family 2 member and flanking sequences are shown in Fig. 3. Several features found in known IS elements are indicated in the figure. Inverted terminal repeats are present and overlined with arrows. These inverted repeats are imperfect but are identical in size (25 bp). Direct repeats corresponding to potential target-site duplications of 9 bp are adjacent to the inverted repeats (see boxed regions). The size of this family 2 member, as determined by the number of base pairs between the target-site duplications (i.e., from inverted repeat to inverted repeat), is 1408 bp. Three major ORFs and five minor ones are present in these 1408 bp. The longest ORF (888 bp) could encode a protein of 296 amino acids and is present on the opposite strand (indicated as ORF 1 in Fig. 3). A second ORF of 486 bp (encoding 162 amino acids) is present on the strand shown and is labeled ORF 2. A third ORF of 480 bp (encoding 160 amino acids) is present on the same strand as ORF 1 and is labeled ORF 3. ORF 3 is contained within ORF 1 but is in a different reading frame. The similarities between the structure of this family 2 member and known IS elements lead us to propose the name IS*T2* for this family of repeated sequences in *T. ferrooxidans*.

Comparison of the Ends of Two Examples of IS*T2*. The ends of a second example of IS*T2* were analyzed and compared to IS*T2* on pTf32. The recombinant plasmid pTf34 (5) has been shown to contain a family 2 member (and therefore IS*T2*). Two fragments from pTf34, each containing a different end of IS*T2*, were selected for further analysis. One end of the family 2 member on the recombinant plasmid pTf34 (see Fig. 1) was subcloned as follows. pTf34 was digested with *Cla* I and *Xho* I, and the resultant fragments were ligated into similarly digested M13. Several plaques were picked at random and used to produce DNAs for sequencing and restriction enzyme analysis. A recombinant M13 phage containing a *Cla* I–*Xho* I insert of approximately 600 bp was selected for further analysis. Nucleotide sequencing of single-stranded template produced by this recombinant phage revealed the presence of sequences similar to those of IS*T2*. The other end of the family 2 member on pTf34 was subcloned by (*i*) cutting pTf34 with *Kpn* I and *Eco*RV, (*ii*) isolating a 900-bp fragment from an agarose gel as described in *Materials and Methods*, (*iii*) further digesting this fragment with *Rsa* I, and (*iv*) ligating the resultant fragments into *Sma* I-digested M13. Single-stranded templates were isolated from several transformants and sequenced to determine which had sequences similar to IS*T2*. Alignment of the sequences of these fragments from pTf34 with the corresponding sequences from pTf32 is shown in Fig. 4. The left end of IS*T2* cloned from pTf32 is identical to a sequence in the fragment cloned from pTf34 (Fig. 4*A*). Directly adjacent to these sequences are target-site duplications of 9 bp (enclosed in the box). The DNA sequences 5′ of the target-site duplications (lowercase letters to the left in Fig. 4*A*) show little similarity, indicating that they probably correspond to unique sequences in the genome. In Fig. 4*B* the sequences corresponding to the other ends of IS*T2* on pTf32 and pTf34 are shown. These sequences are also identical and adjacent to 9-bp target-site duplications (enclosed in the box). The DNA sequences 3′ of the target-site duplications (lowercase letters to the right in Fig. 4*B*) show little similarity, again indicating that they probably correspond to unique sequences. An additional 74

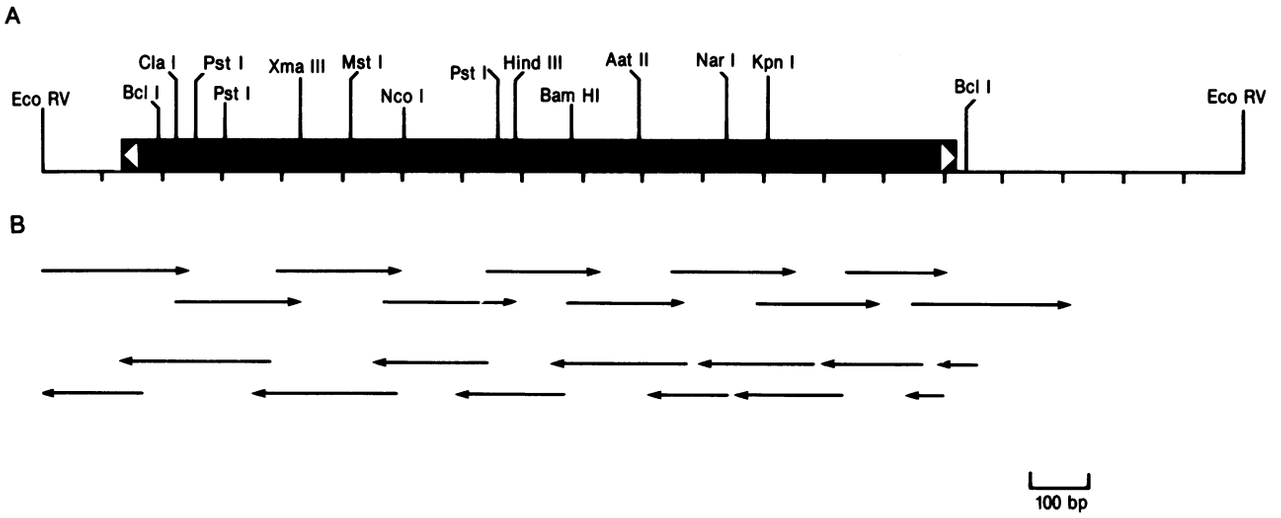

FIG. 2. Restriction map and sequencing strategy of a family 2 member (IS*T2*). (*A*) Partial restriction enzyme site map of a 2.0-kb *Eco*RV–*Eco*RV fragment containing one member of the family 2 repeated sequences of *T. ferrooxidans* (strain ATCC 19859). This fragment was cloned into M13 and sequenced. The thick black line indicates the position of IS*T2* in this fragment. The white triangles at the ends of the thick line indicate the approximate sizes and positions of the terminal inverted repeats. Restriction enzyme sites were inferred from the nucleotide sequence. (*B*) Arrows indicate the extent of sequence determined from overlapping deletions of the fragment shown in *A*. Arrows pointing to the right correspond to sequences of the strand shown in Fig. 3. Those pointing to the left correspond to sequences of the opposite strand. bp, Base pairs.

bases of pTf34 (3′ to the inverted repeat in Fig. 4*A*) were found to be identical to those of pTf32, and 30 bases of pTf34 (5′ to the inverted repeat in Fig. 4*B*) were also identical to those of pTf32 (data not shown), indicating that there is

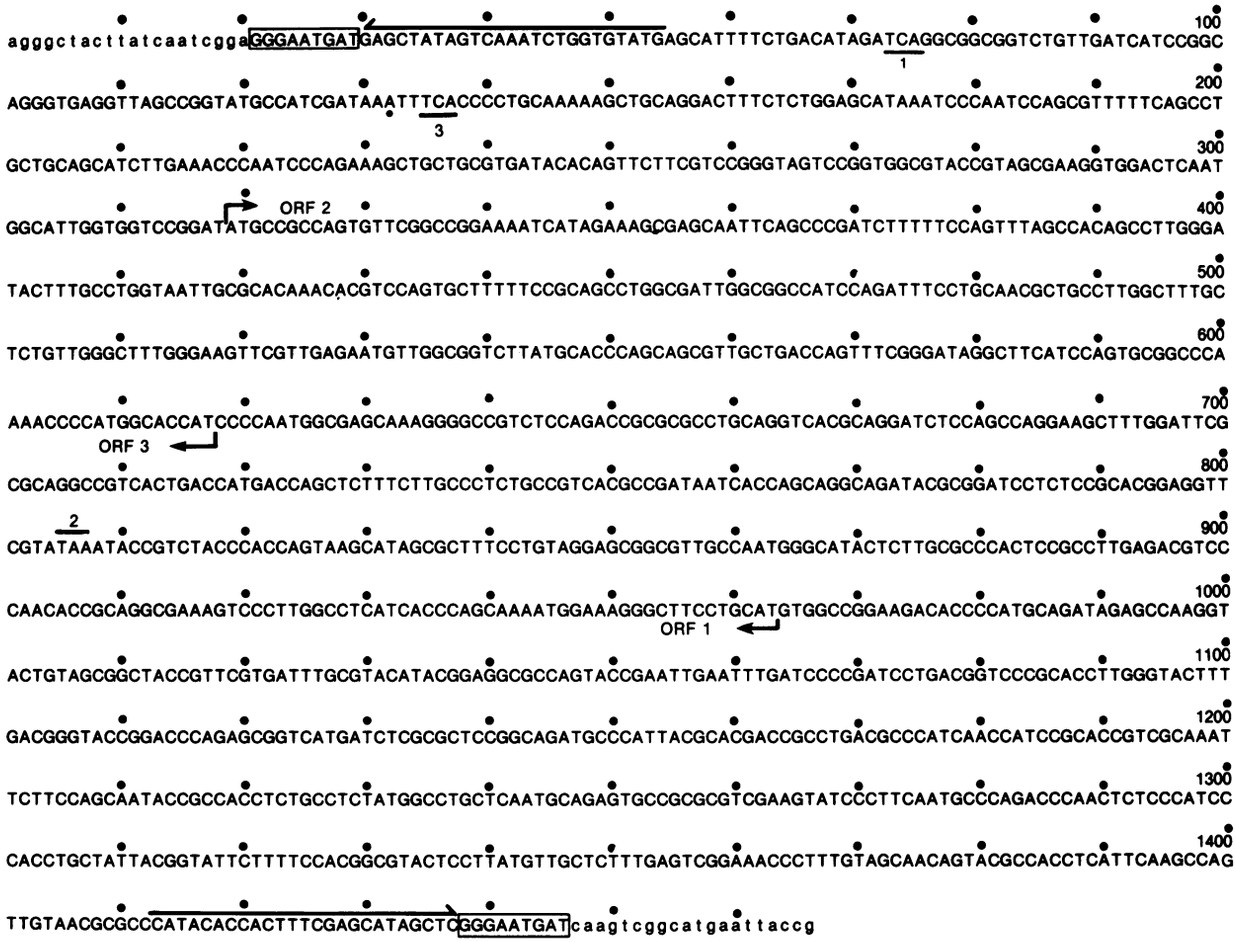

FIG. 3. Nucleotide sequence of IS*T2* and adjacent regions. Lowercase letters indicate unique genomic sequences outside of IS*T2*. Target-site duplications are indicated by uppercase letters enclosed in boxes. The nucleotide sequence of IS*T2* is shown in uppercase letters. The terminal inverted repeats are overlined with arrows and are directly adjacent to the target-site duplications. Locations of the three major ORFs are shown. The initiation codons of ORFs 1 and 3 are marked by left-pointing arrows below the sequence. Since these ORFs are on the opposite strand the reverse complement of ATG (i.e., CAT) is marked. The termination codons for ORFs 1 and 3 are underlined and labeled 1 and 3. The initiation codon of ORF 2 is indicated by a right-pointing arrow above the sequence. The termination codon for this ORF is overlined and labeled 2.

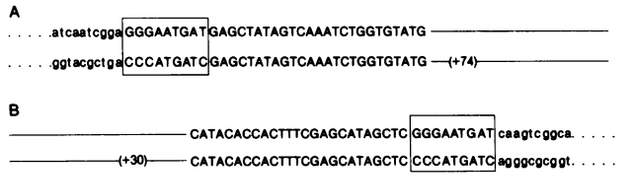

FIG. 4. Comparison of the ends of two examples of IS*T2*. (*A*) The sequences of the "left ends" (compared to Fig. 3) of IS*T2* on pTf32 (upper sequence) and pTf34 (lower sequence). Uppercase letters correspond to IS*T2*. The left halves of the terminal inverted repeats are shown and are identical. An additional 74 bp of IS*T2* on pTf34 were also determined and found to be identical to those on pTf32. A large box encloses the two target-site duplications adjacent to the left inverted repeats. Lowercase letters represent sequences next to the target-site duplications. (*B*) The sequences of the "right-ends" of IS*T2* on pTf32 (upper sequence) and pTf34 (lower sequence). The right halves of the terminal inverted repeats are shown and are identical. An additional 30 bp of IS*T2* on pTf34 were determined and are identical to those of pTf32. A large box encloses the target-site duplications adjacent to the left inverted repeats. Lowercase letters represent sequences next to the target-site duplications.

strong sequence conservation in these regions of the two members of family 2.

DISCUSSION

This report describes an insertion sequence that is present in the genome of *T. ferrooxidans*. We propose the name IS*T2* (Insertion Sequence, *Thiobacillus*, family 2) for this insertion sequence. IS elements have been found in several species of eubacteria (8) and in the archaebacterium *Halobacterium* (6, 13, 14). To our knowledge, IS*T2* is the first example of an IS element in an autotrophic eubacterium. Southern blot analysis of other *T. ferrooxidans* strains has revealed the presence of genomic sequences similar to IS*T2* (15). The possibility exists that IS elements may play an important (presently undetermined) role in the ability of these microorganisms to exist in their unusual environmental niche.

Several lines of evidence indicate that *T. ferrooxidans* has IS elements. First, the presence of two families of repeated sequences in *T. ferrooxidans* has been shown by solution hybridization data as well as Southern blots (5). Second, the nucleotide sequence of one member of family 2 reveals the presence of several features commonly associated with IS elements: inverted terminal repeats, target-site duplications, and three ORFs (Fig. 3). These structural features have been used to identify IS elements in *Shigella sonnei* (16), *E. coli* (17), and *Halobacterium volcanii* (14) in the absence of direct evidence for transposition. Finally, analysis of the genomes of clonal isolates of *T. ferrooxidans* reveals differences in the patterns of bands produced on Southern blots by hybridization probes specific for either family 1 or 2 (15). Preliminary analysis of these data indicate that the band changes are consistent with simple transposition events.

We compared 154 bp of the IS*T2* element of pTf34 with IS*T2* on pTf32 and found that they are identical (Fig. 4). The two regions of IS*T2* on pTf34 examined include the terminal inverted repeats and adjacent sequences. These data, in conjunction with the previously mentioned conservation of restriction sites in these regions of pTf32 and pTf34 (5), reaffirm our conjecture that the members of family 2 exhibit strong sequence conservation.

The target-site duplications and adjacent unique sequences on pTf32 and pTf34 show little sequence similarity (Fig. 4), indicating that IS*T2* does not have rigid target-site specificity. As is the case with other IS elements, the target-site duplications created by both examples of IS*T2* are the same size.

We searched the GenBank Genetic Sequence Data Bank for sequences similar to IS*T2*.¶ Using a program that employed the algorithm of Wilbur and Lipman (18), we found no examples of strong sequence similarity between IS*T2* and the entries in the bacterial section of the data bank (data not shown). We also compared the sequence of IS*T2* with the published sequences of 22 IS elements by using the algorithm of Smith and Waterman (19). No regions of extensive similarity were detected (data not shown). Finally, a comparison of the terminal inverted repeats of IS*T2* to those of the IS elements revealed no extensive similarities.

Investigations into the roles played by IS*T2* and family 1, the other known family of repeated sequences in *T. ferrooxidans*, may further our understanding of repeated DNA sequences in prokaryotes and indicate whether these repeated sequences offer some selective advantage to an organism.

¶EMBL/GenBank Genetic Sequence Database (1987) GenBank (IntelliGenetics, Mountain View, CA), Tape Release 52.0.

We thank Dr. Joseph J. Salvo for his help with DNA sequencing and useful discussions. This work was supported in part by Grant MSM-8215220 from the National Science Foundation to D.S.H.

1. Sapienza, C. & Doolittle, W. F. (1982) *Nature (London)* **295,** 384–389.
2. Schremf, H. (1982) *J. Bacteriol.* **151,** 701–707.
3. Szabo, L. J. & Mills, D. (1984) *J. Bacteriol.* **157,** 821–827.
4. Kaluza, K., Hahn, M. & Hennecke, H. (1985) *J. Bacteriol.* **162,** 535–542.
5. Yates, J. R. & Holmes, D. S. (1987) *J. Bacteriol.* **169,** 1861–1870.
6. Xu, W.-L. & Doolittle, W. F. (1983) *Nucleic Acids Res.* **11,** 4195–4199.
7. Yamada, T., Lee, P.-D. & Kosuge, T. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8263–8267.
8. Iida, S., Meyer, J. & Arber, W. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J. (Academic, Orlando, FL), pp. 159–221.
9. Vishniac, W. V. (1974) in *Bergey's Manual of Determinative Bacteriology*, eds. Buchanan, R. E. & Gibbons, N. E. (Williams & Wilkins, Baltimore), 8th Ed., pp. 456–461.
10. Brierly, C. (1978) *CRC Crit. Rev. Microbiol.* **6,** 207–262.
11. Holmes, D. S. (1984) *Biotechniques* **2,** 68–69.
12. Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* **94,** 441–448.
13. Simsek, M., DasSarma, S., RajBhandary, U. & Khorana, H. G. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 7268–7272.
14. Hofman, J. D., Schalkwyk, L. C. & Doolittle, W. F. (1986) *Nucleic Acids Res.* **14,** 6983–7000.
15. Holmes, D. S., Yates, J. R. & Schrader, J. (1988) *Sci. Tech. Lett.*, in press.
16. Matsutani, S., Ohtsubo, H., Maeda, Y. & Ohtsubo, E. (1987) *J. Mol. Biol.* **196,** 445–455.
17. Kothary, R. K., Jones, D. & Candido, E. P. M. (1985) *J. Bacteriol.* **164,** 957–959.
18. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
19. Smith, T. F. & Waterman, M. S. (1981) *Adv. Appl. Math.* **2,** 482–489.